RoSA Dataset: Road construct zone Segmentation for Autonomous Driving

JINWOO KIM*[1,2], Kyounghwan An[1], and Donghwan Lee[2]

[1] Superintelligence Creative Research Laboratory, ETRI, Korea
[2] Department of Electrical and Engineering, KAIST, Korea
[3] jwkim81@etri.re.kr
[4] https://github.com/jwkim81-ETRI/RoSA-Dataset

Abstract. Current research on road construction environment perception primarily focuses on the detection of objects and signs indicating roadwork. However, this approach requires an additional cognitive step for drivers to fully recognize the extent of construction areas, complicating immediate recognition, especially on highways. Identifying the start of construction zones from a distance is crucial for safe and flexible vehicle rerouting. Existing object detection methods face challenges in identifying these zones from afar due to the small size of marker cones, known as lava cones, which are often spaced widely apart. This can lead to navigational issues when vehicles traverse these gaps. To address these limitations, we propose a novel method that segments construction areas in video footage collectively, enabling the detection of continuous zones from a distance. This approach allows vehicles to adjust their driving paths safely and efficiently. We intend to release a subset of these images with corresponding labeling data to contribute to the field.

Keywords: Autonomous Driving · Roadwork zone · Vision language
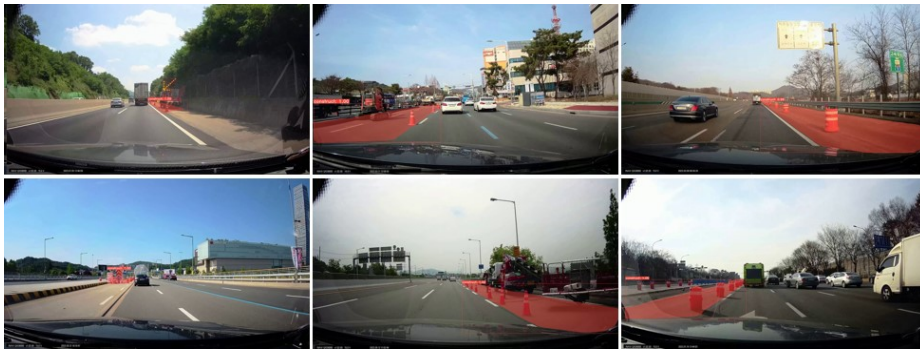
Fig. 1: Road Work Area Segmentation Results in High-Speed Road Environments.

(a) in the distance   **(b)** nearby (cones)   **(c)** nearby (barrel)

1   Introduction

When driving on highways or suburban roads, we often encounter non-standard environments such as roadwork zones. These areas are indicated by various signs, marker cones or barrels, workers, and special vehicles, which drivers recognize in the moment. On expressways, it is crucial for vehicles to detect the start of construction from a sufficient distance, considering the speed limit, to respond safely and flexibly. Objects like marker cones or barrels are small, making them difficult to detect and segment individually, resulting in low accuracy. According to the Korean Road Traffic Act [29], on highways with speeds exceeding 100 km/h, regulations require the installation of flashing lights, signboards, safety barriers, marker cones, signs, and barrels at least 100 meters ahead to ensure that drivers can identify road construction sections. Additionally, for short-term construction, cones are placed at regular intervals, while for long-term projects, barrels are used to clearly demarcate the area.

In this paper, we focus on the rapid and intuitive detection of construction zones in high-speed road environments rather than complex urban areas. Our approach aims to recognize construction areas as they appear in the distance and refine the detection as the vehicle approaches, rather than relying on the detection of individual objects typically associated with road work. Detecting long stretches of road work or construction areas is challenging. However, we intuitively understand the need to identify these zones relative to the ego vehicle and avoid entering them. Therefore, we propose that labeling and detecting construction zones as a single area would be highly efficient. Our dataset includes various scenarios, from long-term highway construction to lane closures on regular roads. The term "construction zone" is used specifically to address the need for early detection of work areas on high-speed roads, as opposed to irregular urban construction sites. We aim to present a data labeling method and its results for faster recognition of construction start points from a distance in highway environments. In construction environments, the spacing between cones and barrels is often significant.

Our perception of these areas as construction zones is based on a continuous temporal understanding of the harmonized region between lane information and the presence of cones and barrels. This paper proposes an efficient data labeling method for segmentation that allows for simultaneous recognition of the construction area and the corresponding lane section. As illustrated in Fig. 1, we show the result of instance segmentation in real time using our provided RoSA dataset. Actually, the configuration of construction zones on multi-lane roads exhibits significant variability depending on their position relative to the ego vehicle. This variability introduces complexity into the detection and classification process. Moreover, in real-world driving scenarios characterized by high traffic volume, the frequent occlusion of construction zone elements by moving objects presents additional challenges to accurate labeling and detection. The interplay between the ego vehicle's position, the layout of the construction zone, and the presence of other road users creates a multifaceted problem space that must be carefully considered in our labeling methodology.

In the subsequent sections, we will detail our strategies for mitigating labeling ambiguities and present guidelines for maintaining consistency in the face of occlusions and varying construction zone configurations. Due to the ambiguous nature of labeling road construction areas in collected video footage, we established specific data labeling definitions. In high-speed road environments, construction zones typically follow a pattern where cones and barrels are placed on lane markings to close off work areas. The United States, China, and Japan, like many other countries, have established traffic regulations and guidelines concerning the placement of cones and barrels in road construction zones [10,30,31]. These standardized practices across various nations support our assertion that the proposed method of segmenting construction areas as unified regions can be efficiently applied. The existence of such regulations underscores the universal recognition of construction zones as distinct, demarcated areas within the road environment.

2   Related Works

Autonomous driving datasets. The KITTI dataset [11] provides a valuable benchmark by offering data acquired from sensors such as cameras, LiDAR, and radar in real road environments. Additionally, the continuous upgrades to datasets like the Waymo Open Dataset [35] and nuScenes [3] over recent years have enabled the capture of a broader array of information from actual road conditions. The expansion of available data, including planning, 3D point cloud, motion, and occupancy data, coupled with ongoing challenges, has significantly enriched the resources available for autonomous driving applications.The Waymo Open Motion Dataset [9] was developed to address the critical need for high-quality motion data; this dataset facilitates the advancement of motion forecasting models essential for autonomous driving planning, especially in interactive scenarios like merges and unprotected turns.

Capturing the challenges of multi-season autonomous driving, the Boreas dataset [2] provides extensive data across diverse weather conditions, enabling benchmarking for odometry, metric localization, and object detection, which are crucial for all-weather, long-term autonomous vehicle navigation. Cityscapes Dataset [7] offers a comprehensive suite for pixel-level, instance-level, and panoptic semantic labeling, supporting the development of vision algorithms that can adapt to complex urban environments. Argoverse [39] expands upon its predecessor and presents a rich collection of datasets with detailed annotations and HD maps, designed to enhance perception and forecasting research in the self-driving domain. As the largest driving video dataset, BDD100K [43] promotes heterogeneous multitask learning with its diverse geographic, environmental, and weather conditions, preparing autonomous driving models for real-world complexity and variability. DAIR-V2X [44] is the first real-world dataset for vehicle-infrastructure cooperative 3D object detection, providing a comprehensive set of data frames, trajectories, vector maps, and traffic lights from natural scenes to advance research in autonomous driving.

Road Obstacle and Workzone Dataset. Chen et al. [4] propose an attention mechanism-based construction zone segmentation method for autonomous driving. The proposed model incorporates spatial and channel attention modules to focus on key features of construction zones, improving segmentation accuracy. Additionally, it introduces a large-scale dataset containing various construction zone scenarios to enhance the model's generalization capabilities. Neuhold et al. [32] present the Mapillary Vistas dataset, offering a global perspective on street scenes with 25,000 high-resolution images annotated with 66 object classes. This dataset is particularly valuable for research on construction zone and road obstacle detection, providing detailed annotations across diverse geographical and cultural contexts. Its comprehensive coverage enhances the development of robust segmentation algorithms for complex urban environments. Wang et al. [37] provide the ApolloScape dataset, a comprehensive resource for autonomous driving research, featuring diverse driving scenarios applicable to construction zone and road obstacle detection. It offers 3D point clouds, high-resolution images, and precise semantic annotations. This multi-modal approach enables the development of sophisticated algorithms capable of understanding complex road environments, including temporary structures and dynamic obstacles. The ROADWork [13] dataset provides extensively processed data for object-based segmentation of road construction elements, including signs, cones, barrels, workers, and specialized vehicles. This comprehensive dataset notably includes drivable path information within construction zones, offering a rich and diverse collection of data to enhance recognition, observation, and analysis capabilities in complex road environments.

Road Obstacle and Workzone Detection. Objects on the road that are not vehicles or means of transportation are typically classified as fallen objects, and the presence of animals requires detection as obstacles. TOYOTA has proposed a method for detecting obstacles on the road by segmenting images based on an auto-encoder [33], creating an anomaly map, and generating an obstacle score map to identify obstructions effectively. Structures to mark construction sections or workspaces, lines to denote and differentiate areas, cones, barriers, and the like for perception were subjects of research even a decade ago [14, 28, 40]. However, detecting specific zones was not easy. Kim et al. [22] propose a method for detecting construction areas in road environments using deep learning. Lee et al. [25] enhanced object detection performance by implementing noise reduction and 3D object recognition techniques in complex environments. Experimental results show high accuracy across various road conditions, demonstrating potential for improving autonomous vehicle safety. Wang et al. [38] introduce RoadWorks-Net, a real-time semantic segmentation network for road construction zone detection. This paper proposes a lightweight encoder-decoder structure, achieving high accuracy while enabling real-time processing. The authors also created a construction zone-specific dataset to improve model performance. Experimental results demonstrate excellent performance across various road and weather conditions.

Autonomous driving Simulator. The OPV2V [42] framework is capable of generating data from various viewpoints within an autonomous driving environment, while V2X-Sim [27] includes imagery that matches point cloud data from different vehicles, both utilizing the CARLA simulator. The dataset for autonomous driving, generated using the CARLA-based simulator [8], is insufficient in providing a diverse and realistic representation of multi-view perspectives of vehicles and potential edge cases that may occur on the road. The simulator's current apabilities do not adequately mimic the complexity and variability of real-world scenarios, which is crucial for the development and testing of robust autonomous driving systems. The Motion Diversification Networks [21] propose a method to reduce the Sim2Real gap by implementing a network in 360-degree videos on the CARLA simulator, which generates diverse and natural motions for pedestrians. This approach enhances the realism of simulated pedestrian behaviors, bridging the gap between virtual simulations and real-world applications in autonomous driving research. GAIA-1 [16], introduced by Wayve, is a generative world model that utilizes video, text, and action inputs to create realistic driving scenarios. It offers granular control over the ego-vehicle behavior and scene features, aiming to enhance the training of autonomous systems.

Segmentation for Autonomous driving. Mask2Former [5] presents a unified architecture for diverse image segmentation tasks using masked attention, while Pointly-Supervised Instance Segmentation [6] introduces an efficient point-based annotation strategy, both enhancing construction zone segmentation. Simple Copy-Paste [12] offers data augmentation through random object pasting, and DAFormer [15] addresses domain adaptation using Transformer encoders and context-aware decoders, improving segmentation in varied construction conditions. Guided Curriculum Model Adaptation [34] adapts daytime models to nighttime environments, crucial for low-light construction areas. Open-Vocabulary Panoptic Segmentation [41] and OpenSeeD [45] advance open-vocabulary segmentation, enabling flexible recognition of diverse construction elements. Li et al. [26] and Zhou et al. [46] propose models for road line detection and real-time road segmentation in complex environments. Kong et al. [24], An et al. [1] and Tao et al. [36] introduce attention-based networks and also Kang ea al. [20] propose the dataset for road detection and semantic segmentation, effectively capturing non-standard road elements in construction areas.

These techniques collectively enhance construction zone segmentation accuracy and efficiency, improving perception capabilities and safety of autonomous vehicles in dynamic road environments. YOLOv8 [18] offers an advanced architecture for object detection and segmentation, potentially enhancing construction zone element detection. YOLOv8-World [19] extends this with global features, suitable for diverse geographical contexts. Meta's SAM [23] provides flexible, prompt-based segmentation, adaptable to complex construction zones without retraining. These techniques collectively enhance the accuracy and efficiency of construction zone segmentation, improving perception capabilities and safety of autonomous vehicles in dynamic road environments.

3   Empirical Data Collection for Road Construction Zone Characterization

The labeling strategy we present takes into account the spatial and temporal aspects of construction zone perception, reflecting the gradual reveal of these areas as a vehicle approaches. This approach not only enhances the accuracy of detection but also provides a more contextually rich understanding of the road environment, potentially improving the safety and efficiency of navigation through these complex scenarios.

In the following Fig. 2, we will detail the specific types of construction zone data included in our dataset, the criteria used for delineating these zones, and the precise methodology employed in our segmentation labeling process. This comprehensive explanation will provide a clear understanding of our data preparation approach and its potential applications in autonomous driving systems.

Our approach aims to align the perception of construction zones in autonomous driving systems with human intuition in everyday driving scenarios. To achieve this, we have defined construction zones in consideration of established traffic regulations, focusing on areas of the road where lanes or parts of lanes are occupied by cones and barrels. This definition seeks to bridge the gap between human cognitive processes and machine perception in the context of road construction areas. Long-term road construction persists regardless of weather conditions, necessitating the recognition of construction zones even during rainfall. Consequently, this study incorporated construction zone data under rainy conditions to ensure the comprehensiveness of the dataset and to reflect the diversity of real driving environments. This inclusion is expected to enhance future performance in construction zone recognition and serve as a crucial factor in guaranteeing the performance of autonomous driving systems across various weather conditions.
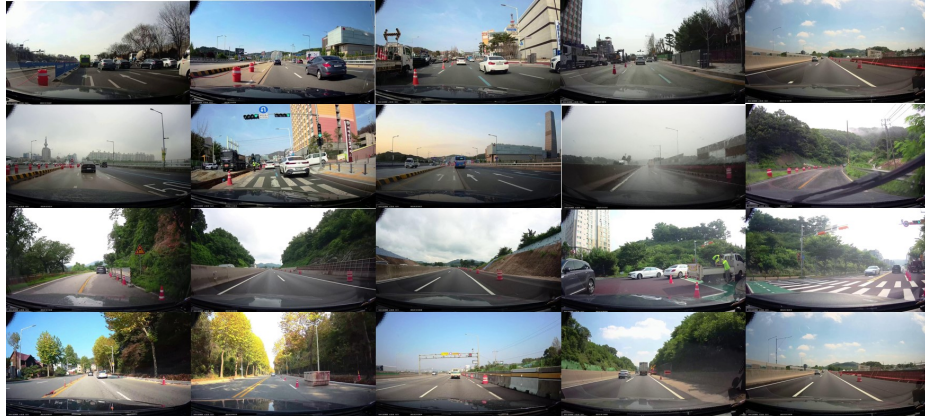
Fig. 2: Diverse road construction and work zones in diverse road environments actually collected, including city roads, highways, and conditions with rain weather

3.1   Data Annotation Methodology and Standardization Criteria

By establishing criteria for data on construction zones in real road environments, we aim to create a labeling methodology that is both precise enough for machine learning applications and intuitive enough to reflect human perception. This approach facilitates the development of autonomous driving systems that can interpret and navigate construction zones in a manner consistent with human drivers' expectations and behavior. We define four primary rules for annotating road construction zones, each addressing specific scenarios commonly encountered on roads, as shown in Figure 3.
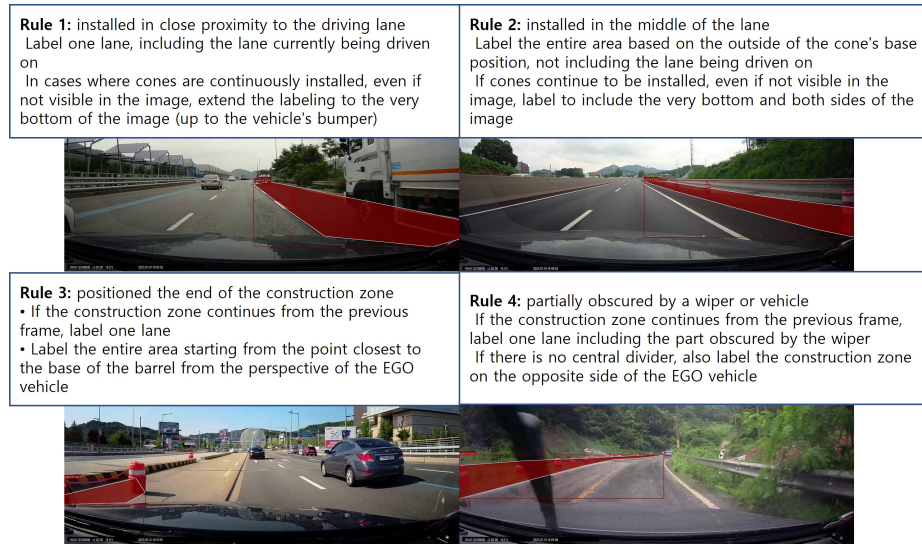
Fig. 3: Standardized annotation rules for diverse road construction zone scenarios: Systematic labeling criteria

Rule 1: Cones in Close Proximity to Driving Lane. When traffic cones are installed close to the active driving lane, we label one full lane width, including the lane currently being driven on. This approach ensures that the immediate danger zone is fully captured. In cases where cones extend beyond the visible frame, the labeling is extended to the bottom edge of the image, typically corresponding to the vehicle's bumper. This rule accounts for the continuity of the construction zone even when visual cues are temporarily obscured.

Rule 2: Cones in the Middle of the Lane. For scenarios where cones are positioned in the center of a lane, we label the entire area based on the outside edge of the cone's base. This labeling explicitly excludes the lane being driven on, focusing on the actual construction area. When the cone arrangement extends beyond the visible frame, the annotation includes the very bottom and both

lateral edges of the image. This comprehensive approach ensures that the full extent of the construction zone is captured, even when partially out of view.

Rule 3: End of Construction Zone. When dealing with the termination of a construction zone, indicated by a single barrel or cone, we employ a context-aware approach. If the construction zone continues from the previous frame, we label one full lane width. The labeling starts from the point closest to the base of the barrel from the ego vehicle's perspective. This rule ensures that the algorithm maintains awareness of the construction zone even as it ends, providing crucial information for safe navigation.

Rule 4: Partial Obstruction Scenarios. In cases where construction markers are partially obscured by windshield wipers or other vehicles, we maintain labeling continuity. If the construction zone persists from the previous frame, we label one lane, including the obscured areas. This approach preserves spatial context despite temporary visual obstructions. Additionally, in the absence of a central divider, we extend the labeling to include construction zones on the opposite side of the ego vehicle, ensuring comprehensive scene understanding.

In the following sections, we will elaborate on the practical application of these criteria in our labeling process, addressing challenges such as partial occlusions, varying lighting conditions, and the dynamic nature of construction zones. This detailed exposition will provide a comprehensive understanding of our approach to creating a robust dataset for training and evaluating construction zone detection algorithms in autonomous driving systems.

3.2   Definition of Road Construction Zone Data Labeling Methodology

Occlusions caused by other vehicles can intermittently obscure key visual cues that denote the presence and extent of a construction zone. This phenomenon introduces ambiguity into the labeling process, as the true boundaries and characteristics of the construction area may not be consistently visible across all frames of a video sequence or in isolated images.

Fig. 4: Segmentation results of road construction zones with and without including obstacles within the zone. **Top:** Labeling based on road surface using green line as reference. **Bottom:** Labeling including various objects in construction zone; yellow boxes show reduced accuracy.

To address these challenges, our labeling approach must account for:

1. Spatial variability: The diverse configurations of construction zones relative to the ego vehicle's position and trajectory.
2. Temporal consistency: The need for coherent labeling across video frames, even when key elements are temporarily obscured.
3. Partial visibility: The ability to infer the presence and extent of construction zones from limited visual cues.
4. Contextual information: The integration of broader scene understanding to support accurate labeling in ambiguous situations.

By acknowledging and systematically addressing these complexities, we aim to develop a labeling methodology that captures the nuanced reality of construction zones in diverse traffic conditions. This approach will provide a more reliable foundation for training detection algorithms, ultimately enhancing the ability of autonomous vehicles to navigate these challenging road environments safely and efficiently.

Our approach to labeling construction zones excludes objects within the zone such as construction vehicles, personnel, structures, or special vehicles, beyond the cones and barrels. As illustrated in the segmentation results in Figure 4, we implemented a ground-level cut-off for labeling, represented by the green line in the upper row of images. This decision was made based on performance considerations. As shown in the yellow boxed areas of the lower images in Figure 4, including irregular objects above ground level led to significant irregularities in the overall shape of the construction zone, resulting in substantial performance degradation. This labeling strategy aims to maintain consistency and improve the overall performance of construction zone detection algorithms by focusing on the fundamental ground-level demarcation of these areas. To acquire data on construction zones occasionally encountered during road driving, we utilized a commercial vehicle dashcam (manufactured by INAVI [17] in Figure 5) to capture videos of various scenarios, which were then reconstructed into sequential images. The RoSA dataset images were captured with FHD (1920x1080) resolution, a 155-degree horizontal field of view (HOV), and HDR capability.

(a) Black box   (b) Data acquisition vehicle   (c) Our Autonomous Vehicle
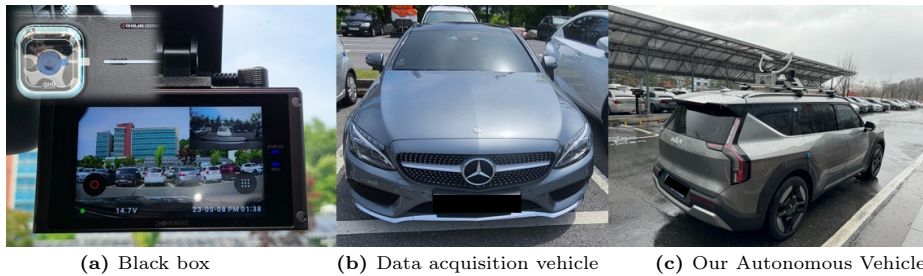
**Fig. 5:** Data acquisition environment.

4   Analysis and Experiments

4.1   Roadworks Zone Instance Segmentation

RoSA Dataset for Training. The dataset encompasses 15 distinct road scenario types, including both highway (8 types) and urban road (7 types) environments, featuring construction zones marked by barrels (8 types) and cones (7 types). The RoSA dataset comprises a total of 2,664 frames, with 2,131 frames allocated for training and 722 frames for validation, maintaining an 8:2 ratio. To rigorously evaluate the generalization capability of models trained on our dataset, we structured our validation set to include a significant portion of novel scenarios. Specifically, 25% of the validation set, comprising 186 images, contains entirely different scenarios from those present in the training data. This deliberate inclusion of unseen configurations in the validation set serves to assess the robustness and adaptability of the trained models to new and potentially challenging construction zone layouts. Validation data was extracted at uniform intervals from the entire image sequence to ensure comprehensive representation. This diverse and well-structured dataset aims to provide a robust foundation for developing and evaluating algorithms for construction zone detection and segmentation in various road conditions, contributing to the advancement of autonomous driving technologies.

Segmentation. We present the temporal progression of instance segmentation results utilizing our construction zone dataset. Figure 6a illustrates the detection outcome of a cone-delineated construction area in an urban setting. The detection algorithm integrates the recognition of cones with lane demarcations. Notably, the system demonstrates persistent identification of the continuous construction zone, even when the proximal cones have passed beyond the camera's field of view, showcasing the robustness of the approach in maintaining spatial context. Similarly, the results in Figure 6b demonstrate the detection of construction zones in a highway environment. As observed in the last two frames, despite the presence of only a single cone, the system successfully detects the area as a continuous region, maintaining connectivity with the previous frames. This exemplifies the algorithm's ability to preserve spatial and temporal continuity in the detection process, even with limited visual cues.

Figure 7 depicts the detection of long-term construction zones delineated by barrels. Specifically, Figure 7a illustrates the detection results of barrel-defined areas in an urban environment, while Figure 7b demonstrates the detection of extended construction zones marked by barrels in a highway setting. Consistent with the findings presented in Figure 6 and Figure 7, these results showcase the system's capability to maintain continuous detection of the construction zone, even when individual barrels move out of the camera's field of view. This persistent detection underscores the robustness of our approach in maintaining spatial and temporal continuity across diverse road environments and construction zone configurations.

(a) City road construction with cones

(b) Highway road construction with cones
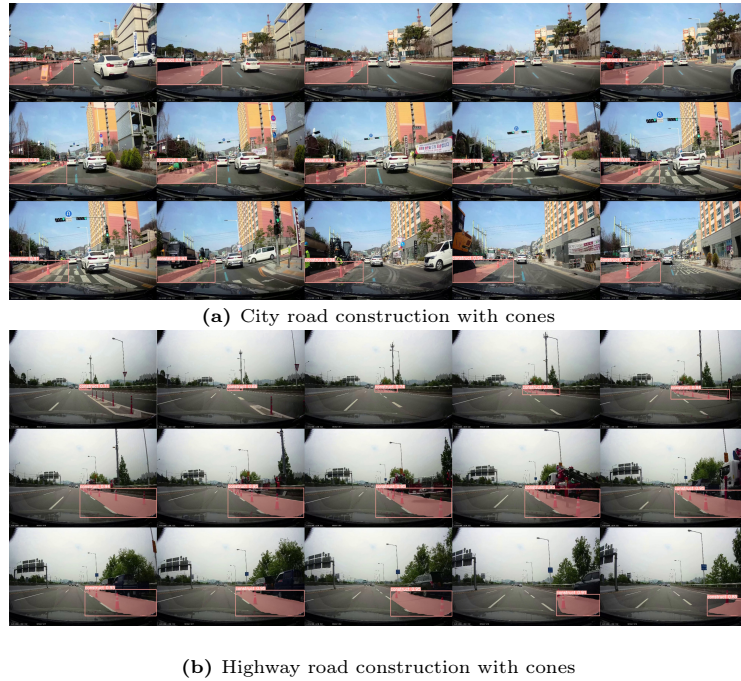
Fig. 6: The segmentation of cones-based zones in real roads.

4.2   Segmentation Performance Results Based on RoSA Dataset

We evaluated the applicability of instance segmentation to the RoSA dataset by examining various open-source methodologies. We selected YOLOv8 [18] as our benchmark model due to its recent superior performance. By applying a range of model architectures from lightweight to high-performance, we sought to comprehensively assess the practicality of the RoSA dataset.

Our proposed performance evaluation criterion for construction zone segmentation primarily focused on the accuracy of detecting the nearest construction zone boundary relative to the ego vehicle. Moreover, considering the real-time operational requirements of autonomous vehicles, we aimed to achieve excellent performance while utilizing lightweight models. Experimental results demonstrated that even with relatively lightweight models, the segmentation detection performance based on the RoSA dataset was competitive compared to the COCO dataset-based YOLOv8 [18]. This finding validates the feasibility of not only individual object-based segmentation but also the area-based detection method for construction zones proposed in this study.

The primary objective of this research was to verify the potential of instance segmentation using the RoSA dataset. We anticipate further performance improvements with advancements in network architectures, progress in generative AI technologies, and expansion of the dataset. These research outcomes are ex-

(a) City road construction with barrel

(b) Highway road construction with barrel
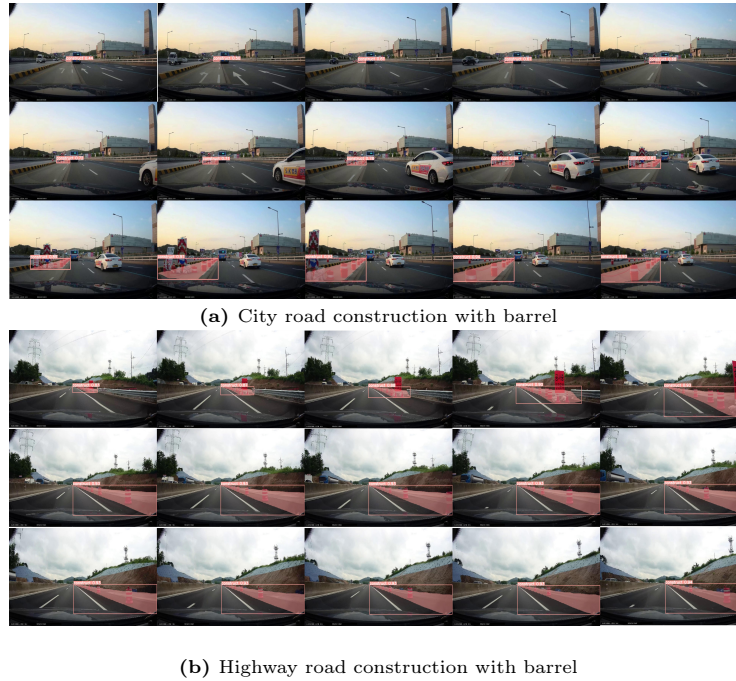
Fig. 7: The segmentation of barrel-based zones in real roads.

pected to contribute significantly to the development of construction zone recognition technology in autonomous driving environments.

The presentation of the RoSA dataset's performance using YOLOv8 [18] serves primarily as an illustrative benchmark for potential users, leveraging a methodology widely recognized for its usability and effectiveness. This approach is intended to demonstrate the dataset's applicability rather than to propose a definitive solution. We explicitly state that our actual vehicle system will employ alternative algorithms optimized for our specific requirements and constraints. This strategy not only validates the utility of the RoSA dataset using a popular, high-performance model but also establishes a baseline for comparison across various segmentation methodologies. By doing so, we enable researchers and developers to assess the dataset's potential and versatility, while acknowledging that optimal performance in real-world applications may necessitate custom-designed algorithms that capitalize on the unique characteristics of the RoSA dataset and address the specific challenges of construction zone segmentation in autonomous driving scenarios. Table 1 presents the performance of five YOLOv8 variants fine-tuned on the RoSA dataset, tested using an RTX 4090 GPU. Results demonstrate that lightweight networks can effectively detect construction zones, indicating their potential for real-time autonomous driving applications while balancing model complexity and accuracy.

Table 1: YOLOv8 Segmentation Models based on RoSA Performance

Model	Size (pixels)	mAP^{box} 50-95	mAP^{mask} 50-95	Speed (ms) (RTX 4090 TensorRT)
YOLOv8n-seg-RoSA	640	89.1	79.6	3.1
YOLOv8s-seg-RoSA	640	90.3	81.9	5.1
YOLOv8m-seg-RoSA	640	91.3	85.1	9.4
YOLOv8l-seg-RoSA	640	90.1	82.4	14.1
YOLOv8x-seg-RoSA	640	88.6	80.8	20.3

5 Discussion

Future Plans The data presented in this paper were collected to detect construction zones frequently encountered in Korean road environments. Data were stored whenever diverse construction scenarios occurred. To facilitate continuous data collection, we utilized a dashcam capable of storing data upon artificial impact. We could not find publicly available labeled datasets of construction zone images collected in high-speed driving environments similar to the data we present. Most existing datasets primarily focus on object detection rather than comprehensive zone segmentation. Moreover, our proposed format and labeling method are simple and intuitive, distinguishing them from existing object-based labeled data. The approach presented in this paper labels construction zone areas as single image units while also considering the relationship between consecutive frames to partially predict areas.

In the future, we intend to collect and process additional data on construction zone environments under various weather conditions, albeit rare, and make them public. Furthermore, we will add prompt data to enable continuous situation awareness using advanced methods such as LLMs (Large Language Models) or VLMs (Vision-Language Models). This research contributes to the field by providing a unique dataset of construction zones in high-speed driving environments, with a novel labeling approach that considers both individual frames and temporal continuity. Our commitment to expanding the dataset with diverse environmental conditions and incorporating advanced AI techniques demonstrates the potential for significant advancements in autonomous driving technology, particularly in the challenging domain of construction zone detection.

To facilitate real-time recognition system implementation, our research team is exploring the direct installation of cameras or dashcam sensors with identical field of view and resolution on autonomous vehicles. In subsequent research, we intend to present results from the application of models trained on this dataset, supplemented with additional road data and synthetic data, to actual autonomous vehicles. This approach aims to validate the practicality and efficacy of our research, potentially contributing significantly to the advancement of autonomous driving technology.

Limitations The dataset and labeling methodology presented in this study have certain inherent limitations. For instance, ambiguity may arise when attempting to determine the continuity of construction zones based solely on single images. While our dataset is composed of individual images, each scenario comprises image sequences with temporal continuity. We posit that these limitations can be mitigated by leveraging the scenario-based structure of our dataset to apply video segmentation techniques. The immediate application of this dataset presents compatibility challenges for autonomous vehicles with diverse camera sensor configurations. However, as our data collection utilized a standard dashcam, we anticipate that customized implementation for various vehicles can be achieved by installing cameras with comparable specifications.

6   Conclusion

We have presented a novel data structure and labeling methodology for segmenting construction zones encountered on both regular roads and highways. Our approach encompasses a diverse range of environmental conditions, providing a comprehensive dataset for the development and evaluation of autonomous driving systems. The primary focus of our labeling strategy is to enable rapid and intuitive recognition of construction areas, particularly crucial for high-speed driving scenarios. By treating each construction zone as a single instance object, our labeling method maintains robustness against momentary occlusions caused by other vehicles or environmental factors such as windshield wipers, ensuring consistent detection capabilities. Recent advancements in segmentation techniques, especially those leveraging vision-language models, have demonstrated remarkable performance improvements. We anticipate that the unique segmentation dataset proposed in this paper will benefit significantly from these cutting-edge approaches and future developments in the field. Our goal is to catalyze further research and innovation in this critical area of autonomous driving technology. By making these data publicly available, we aim to contribute substantially to the development of robust construction zone recognition technologies for high-speed autonomous driving applications. This ongoing effort addresses one of the most frequent and challenging unstructured environments encountered in autonomous driving scenarios. In conclusion, our work provides a foundation for more accurate and efficient detection of construction zones. We believe that this dataset and labeling approach will play a vital role in advancing the state-of-the-art in this domain, ultimately contributing to the broader goal of safer and more capable autonomous driving systems.

Acknowledgements

References

1. An, T., Kang, J., Choi, D., Min, K.W.: Crfnet: Context refinement network used for semantic segmentation. ETRI Journal (October 2023). `https://doi.org/10.4218/etrij.2023-0017`

2. Burnett, K., Yoon, D.J., Wu, Y., Li, A.Z., Zhang, H., Lu, S., Qian, J., et al.: Boreas: A multi-season autonomous driving dataset. International Journal of Robotics Research (IJRR) (2023)

3. Caesar, H., Bankiti, V., Lang, A.H., Vora, S., Liong, V.E., Xu, Q., Krishnan, A., Pan, Y., Baldan, G., Beijbom, O.: nuscenes: A multimodal dataset for autonomous driving. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 11621–11631 (2020)

4. Chen, L., Wu, J., Liu, R.: Attention-guided construction zone segmentation for autonomous driving. In: Asian Conference on Computer Vision (2021)

5. Cheng, B., Misra, I., Schwing, A.G., Kirillov, A., Girdhar, R.: Masked-attention mask transformer for universal image segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2022)

6. Cheng, B., Parkhi, O., Kirillov, A.: Pointly-supervised instance segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2022)

7. Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., Schiele, B.: The cityscapes dataset for semantic urban scene understanding. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016)

8. Dosovitskiy, A., Ros, G., Codevilla, F., Lopez, A., Koltun, V.: Carla: An open urban driving simulator. In: Proceedings of the 1st Annual Conference on Robot Learning. pp. 1–16 (2017)

9. Ettinger, S., Cheng, S., Caine, B., Liu, C., Zhao, H., Pradhan, S., Chai, Y., Sapp, B., Qi, C.R., Zhou, Y., et al.: Large scale interactive motion forecasting for autonomous driving: The waymo open motion dataset. In: Proceedings of the International Conference on Computer Vision (ICCV) (2021)

10. Federal Highway Administration: Manual on uniform traffic control devices for streets and highways. Official publication, U.S. Department of Transportation, Washington, D.C. (2009), available at: `https://mutcd.fhwa.dot.gov/pdfs/2009r1r2/mutcd2009r1r2edition.pdf`

11. Geiger, A., Lenz, P., Urtasun, R.: Are we ready for autonomous driving? the kitti vision benchmark suite. In: Conference on Computer Vision and Pattern Recognition (CVPR) (2012)

12. Ghiasi, G., Cui, Y., Srinivas, A., Qian, R., Lin, T., Cubuk, E.D., Le, Q.V., Zoph, B.: Simple copy-paste is a strong data augmentation method for instance segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2022)

13. Ghosh, A., Tamburo, R., Zheng, S., Alvarez-Padilla, J.R., Zhu, H., Cardei, M., Dunn, N., Mertz, C., Narasimhan, S.G.: Roadwork dataset: Learning to recognize, observe, analyze and drive through work zones. In: TBD. Carnegie Mellon University (2024), `https://www.cs.cmu.edu/~ILIM/roadwork_dataset`

14. Graf, R., Wimmer, A., Dietmayer, K.C.: Probabilistic estimation of temporary lanes at road work zones. In: Proceedings of the International Conference on Intelligent Transportation Systems (ITSC) (2012)

15. Hoyer, L., Dai, D., Van Gool, L.: Daformer: Improving network architectures and training strategies for domain-adaptive semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2022)

16. Hu, A., Russell, L., Yeo, H., Murez, Z., Fedoseev, G., Kendall, A., Shotton, J., Corrado, G.: Gaia-1: A generative world model for autonomous driving. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2023), `https://arxiv.org/abs/2309.17080`

17. INAVI: Inavi qxd5000 black box. `https://www.inavi.com/Products/BlackBox/Gate?target=_QXD5000` (2024), product Specification

18. Jocher, G., Chaurasia, A., Qiu, J.: Yolov8: A high-performance object detection and segmentation model. arXiv preprint arXiv:2305.09972 (2023)

19. Jocher, G., Chaurasia, A., Qiu, J.: Yolov8-world: A real-time global object detection and segmentation model. arXiv preprint arXiv:2308.05975 (2023)

20. Kang, J., Han, S.J., Kim, N., Min, K.W.: Etli: Efficiently annotated traffic lidar dataset using incremental and suggestive annotation. ETRI Journal (September 2021). `https://doi.org/https://doi.org/10.4218/etrij.2021-0055`

21. Kim, H.J., Ohn-Bar, E.: Motion diversification networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1650–1660 (2024)

22. Kim, J., Park, J., Kim, S.: Detection of construction areas in road environments using deep learning. In: IEEE Transactions on Intelligent Transportation Systems (2019)

23. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., Dollár, P., Girshick, R.: Segment anything. arXiv preprint arXiv:2304.02643 (2023)

24. Kong, H., Audibert, J.Y., Ponce, J.: Seanet: Semantic embedding attention network for road detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2020)

25. Lee, D., Han, S.J., Min, K.W., Choi, J., Park, C.H.: Emos: Enhanced moving object detection and classification via sensor fusion and noise filtering. ETRI Journal (September 2023). `https://doi.org/https://doi.org/10.4218/etrij.2023-0109`

26. Li, X., Li, J., Hu, X., Yang, J.: Line-cnn: End-to-end traffic line detection with line proposal unit. In: IEEE Transactions on Intelligent Transportation Systems (2019)

27. Li, Y., An, Z., Wang, Z., Zhong, Y., Chen, S., Feng, C.: V2x-sim: A virtual collaborative perception dataset for autonomous driving. arXiv preprint arXiv:2202.08449 (2022). `https://doi.org/10.48550/ARXIV.2202.08449`

28. Mathibela, B., Osborne, M.A., Posner, I., Newman, P.: Can priors be trusted? learning to anticipate roadworks. In: Proceedings of the International Conference on Intelligent Transportation Systems (ITSC) (2012)

29. Ministry of Land, Infrastructure and Transport: Guidelines for traffic management in road construction sites. Technical guidelines, Ministry of Land, Infrastructure and Transport, Sejong, South Korea (2022), available at: `http://www.molit.go.kr/USR/I0204/m_45/dtl.jsp?idx=15952`

30. Ministry of Transport of the People's Republic of China: Technical specifications for design and construction of highway traffic safety facilities. Technical standards, Ministry of Transport of the People's Republic of China, Beijing, China (2016), document Number: JTG D81-2017, Available at: `https://www.mot.gov.cn/zhengcejiedu/jsdajtssgf/`

31. National Highway Traffic Safety Administration: Work zone safety. Safety report, U.S. Department of Transportation, Washington, D.C. (2021), available at: `https://www.nhtsa.gov/road-safety/work-zone-safety`
32. Neuhold, G., Ollmann, T., Rota Bulò, S., Kontschieder, P.: The mapillary vistas dataset for semantic understanding of street scenes. In: Proceedings of the IEEE International Conference on Computer Vision (2017)
33. Ohgushi, T., Horiguchi, K., Yamanaka, M. (eds.): Computer Vision – ACCV 2020. Springer (2021). `https://doi.org/10.1007/978-3-030-69544-6_14`
34. Sakaridis, C., Dai, D., Van Gool, L.: Guided curriculum model adaptation and uncertainty-aware evaluation for semantic nighttime image segmentation. In: ICCV (2019)
35. Sun, P., Kretzschmar, H., Dotiwalla, X., Chouard, A., Patnaik, V., Tsui, P., Guo, J., Zhou, Y., Chai, Y., Caine, B., et al.: Scalability in perception for autonomous driving: Waymo open dataset. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2446–2454 (2020)
36. Tao, A., Sapra, K., Catanzaro, B.: Hierarchical multi-scale attention for semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2020)
37. Wang, P., Huang, X., Cheng, X., Zhou, D., Geng, Q., Yang, R.: The apolloscape open dataset for autonomous driving and its application. In: IEEE Transactions on Pattern Analysis and Machine Intelligence (2019)
38. Wang, X., Li, Y., Zhang, H.: Roadworks-net: A real-time semantic segmentation network for road construction zone detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (2020)
39. Wilson, B., Qi, W., Agarwal, T., Lambert, J., Singh, J., Khandelwal, S., Pan, B., Kumar, R., et al.: Argoverse 2: Next generation datasets for self-driving perception and forecasting. arXiv preprint arXiv:2301.00493 (2023)
40. Wimmer, A., Weiss, T., Flogel, F., Dietmayer, K.: Automatic detection and classification of safety barriers in road construction sites using a laser scanner. In: Proceedings of the IEEE Intelligent Vehicles Symposium (2009)
41. Xu, J., Liu, S., Vahdat, A., Byeon, W., Wang, X., De Mello, S.: Open-vocabulary panoptic segmentation with text-to-image diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2023)
42. Xu, R., Xiang, H., Xia, X., Han, X., Li, J., Ma, J.: Opv2v: An open benchmark dataset and fusion pipeline for perception with vehicle-to-vehicle communication. In: IEEE International Conference on Robotics and Automation (ICRA). pp. 2583–2589 (2022)
43. Yu, F., Chen, H., Wang, X., Xian, W., Chen, Y., Liu, F., Madhavan, V., Darrell, T.: Bdd100k: A diverse driving dataset for heterogeneous multitask learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2020)
44. Yu, H., Luo, Y., Shu, M., Huo, Y., Yang, Z., Shi, Y., Guo, Z., Li, H., Hu, X., Yuan, J., et al.: Dair-v2x: A large-scale dataset for vehicle-infrastructure cooperative 3d object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 21361–21370 (2022)
45. Zhang, H., Li, F., Zou, X., Liu, S., Li, C., Yang, J., Zhang, L.: A simple framework for open-vocabulary segmentation and detection. In: ICCV (2023)
46. Zhou, Y., Zhu, H., Liu, Z., Shen, S., Wang, H.: Da-roadseg: Diverse annotation for real-time road segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2020)